

Aplicaciones Científicas de la Programación Paralela en el Grupo de Computación Científica y Programación Paralela de la Universidad de Murcia

Domingo Giménez
`dis.um.es/~domingo`

Grupo de Computación Científica y Programación Paralela
`luna.inf.um.es/grupo_investigacion`

Departamento de Informática y Sistemas, Universidad de Murcia



Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltante de Datos (DEA)
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo
- 6 Créditos

Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltante de Datos (DEA)
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo
- 6 Créditos

Grupo CCPP

Líneas de trabajo

- Aplicaciones
Ejemplo de [Francisco José Herrera](#).
Presentación en página del grupo.
- Computación Numérica
Algoritmos Matriciales Paralelos, y Autooptimización.
Ejemplo de [Jesús Cámara](#).
- Optimización
Metaheurísticas paralelas.
Ejemplo de [José Matías Cutillas](#).

Aplicaciones en el grupo de CCPP

`luna.inf.um.es/grupo_investigacion`

	Numérico	Optimización
Simulación climática	█	
Hidrodinámica	█	
Electromagnetismo	█	█
Representación del terreno	█	█
Tratamiento de señal acústica	█	█
Tratamiento de imágenes	█	
<i>docking</i> de moléculas		█
Econometría	█	
Análisis envolvente de datos	█	█
Constantes cinéticas		█
Consumo electricidad		█
Análisis de sedimentos	█	█
Series temporales	█	█
Sistemas multicuerpo	█	█

Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envolvente de Datos (DEA)**
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo
- 6 Créditos

Problema

- Análisis de eficiencia de varias unidades de decisión (DMU)
- Cada unidad k varias entradas x y salidas y
- Problema de optimización para cada unidad
- Disponibles métodos exactos, tipo Branch and Bound, y librerías (CPLEX)
- alto coste de ejecución
- necesidad de **paralelismo** y métodos aproximados (**metaheurísticas**)
- y operaciones **matriciales** en la evaluación de las restricciones y cálculo del fitness
- Posibilidad de métodos híbridos: exactos con metaheurísticas

Martín González, Jose J. López-Espín, Juan Aparicio, Domingo Giménez, Jesús T. Pastor: Using Genetic Algorithms for Maximizing Technical Efficiency in Data Envelopment Analysis, The International Conference on Computational Science, Reykjavík, Iceland, 1-3 June, 2015

Modelo de Programación Matemática

$$\max \quad \beta_k - \frac{1}{m} \sum_{i=1}^m \frac{t_{ik}^-}{x_{ik}}$$

s.t.

$$\beta_k + \frac{1}{s} \sum_{r=1}^s \frac{t_{rk}^+}{y_{rk}} = 1 \quad (c.1)$$

$$-\beta_k x_{ik} + \sum_{j=1}^n \alpha_{jk} x_{ij} + t_{ik}^- = 0 \quad \forall i \quad (c.2)$$

$$-\beta_k y_{rk} + \sum_{j=1}^n \alpha_{jk} y_{rj} - t_{rk}^+ = 0 \quad \forall r \quad (c.3)$$

$$-\sum_{i=1}^m \nu_{ik} x_{ij} + \sum_{r=1}^s \mu_{rk} y_{rj} + d_{jk} = 0 \quad \forall j \quad (c.4)$$

$$\nu_{ik} \geq 1 \quad \forall i \quad (c.5)$$

$$\mu_{rk} \geq 1 \quad \forall r \quad (c.6)$$

$$d_{jk} \leq M b_{jk} \quad \forall j \quad (c.7)$$

$$\alpha_{jk} \leq M(1 - b_{jk}) \quad \forall j \quad (c.8)$$

$$b_{jk} = 0, 1 \quad \forall j \quad (c.9)$$

$$\beta_k \geq 0 \quad (c.10)$$

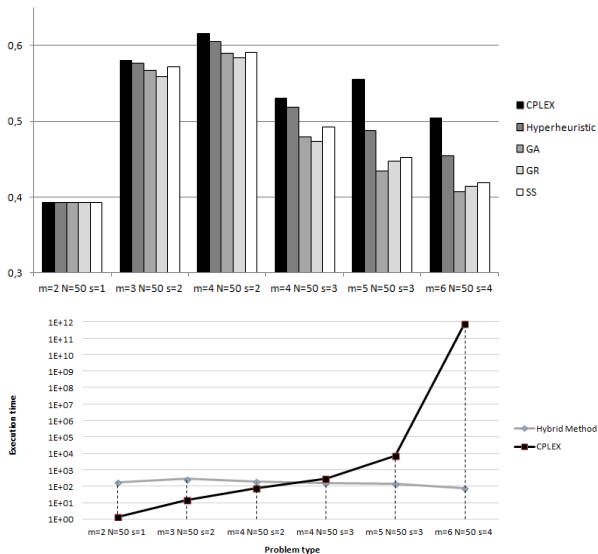
$$t_{ik}^- \geq 0 \quad \forall i \quad (c.11)$$

$$t_{rk}^+ \geq 0 \quad \forall r \quad (c.12)$$

$$d_{jk} \geq 0 \quad \forall j \quad (c.13)$$

$$\alpha_{jk} \geq 0 \quad \forall j \quad (c.14)$$

Resultados



Combinación con métodos numéricos de optimización e hiperheurísticas

Schemes	Hyperheuristic	Metaheuristic
$m = 2, n = 50, s=1$	0.393	0.393
$m = 3, n = 50, s=2$	0.576	0.579
$m = 4, n = 50, s=2$	0.606	0.609
$m = 4, n = 50, s=3$	0.518	0.520
$m = 5, n = 50, s=3$	0.488	0.502
$m = 6, n = 50, s=4$	0.455	0.430

Martin Gonzalez-Rodríguez, Jose-Juan López-Espín, Juan Aparicio, Domingo Giménez, El-Ghazali Talbi: A parameterized scheme of meta-heuristics with exact methods for determining the Principle of Least Action in Data Envelopment Analysis. CEC 2017: 588-595

Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltante de Datos (DEA)
- 3 Series temporales**
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo
- 6 Créditos

Series Temporales Multivariantes

- Con **Big Data**, cada vez más datos de diversidad de campos.
- En algunos casos, tienen evolución temporal: **series temporales**.
- Método tradicional de obtención de modelos por expertos es cada vez de más difícil aplicación.
- **Series Temporales Multivariantes:**
Parámetros de los que se toman valores periódicamente, determinar la evolución temporal de los parámetros, el valor de un parámetro en un instante puede depender de sus valores en instantes anteriores y de los de otros parámetros. Puede haber parámetros externos, que influyen en los internos pero no influidos por estos.

Datos de la serie temporal

- d parámetros, t instantes de tiempo
Serie temporal es $Y \in R^{t \times d}$:

$$\begin{pmatrix} y_1^{(1)} & \cdots & y_d^{(1)} \\ \vdots & \ddots & \vdots \\ y_1^{(t)} & \cdots & y_d^{(t)} \end{pmatrix}$$

- Dependencias temporales con i instantes de tiempo anteriores.
- Datos externos al modelo, z , que influyen en los valores de y pero no se ven influenciados:

$$\begin{pmatrix} z_1^{(1)} & \cdots & z_e^{(1)} \\ \vdots & \ddots & \vdots \\ z_1^{(t)} & \cdots & z_e^{(t)} \end{pmatrix}$$

Formulación matricial

Dependencia de valores de un instante j en función de los valores anteriores:

$$y^{(j)} \approx y^{(j-1)}A_1 + y^{(j-2)}A_2 + \dots + y^{(j-i)}A_i + z^{(j-1)}B_1 + z^{(j-2)}B_2 + \dots + z^{(j-k)}B_k + a$$

A_l , $1 \leq l \leq i$, de dimensión $d \times d$, representa la dependencia de los datos con los valores de l instantes anteriores;

B_l , $1 \leq l \leq k$, de dimensión $e \times d$, dependencias de factores externos.

En forma matricial:

$$Y(i+1:t,:) \approx Y(i:t-1,:)A_1 + Y(i-1:t-2,:)A_2 + \dots + Y(1:t-i,:)A_i + Z(i:t-1,:)B_1 + Z(i-1:t-2,:)B_2 + \dots + Z(i-k+1:t-k,:)B_k + A_0 =$$

$$[Y(i:t-1,)|\dots|Y(1:t-i,)|Z(i:t-1,)|\dots|Z(i-k+1:t-k,)|1] \begin{bmatrix} A_1 \\ \vdots \\ A_i \\ B_1 \\ \vdots \\ B_k \\ a \end{bmatrix}$$

Problema de mínimos cuadrados

- Obtener el modelo consiste en calcular A_l , B_l y a que minimizan la diferencia.
- Problema de mínimos cuadrados con matriz de incógnitas de dimensión $(d \cdot i + e \cdot k + 1) \times d$.
- \hat{Y} es $Y(i + 1 : t, :)$, \hat{X} la matriz formada por submatrices de Y , Z y el vector 1, y \hat{A} la formada por las matrices A_l , B_l y el vector a :

$$\min_{\hat{A}} \left\| \hat{Y} - \hat{X}\hat{A} \right\|$$

Tipos:

- Sin restricciones en las entradas de las matrices.
- Puede que los valores estén en unos ciertos rangos,
- o que sean de un conjunto.
- Puede que los valores de un parámetro no influyan en los de otro.

Problemas matriciales

- Rutinas y librerías para estos tipos de problemas (Matlab, LAPACK).
- Métodos basados en descomposiciones matriciales y métodos iterativos.
- Iterativos: partir de valores proporcionados por los expertos en el campo de los datos, o resolver ecuaciones parciales y comenzar la búsqueda desde ellas.
- Explotación de la estructura tipo Toeplitz por vectores.
- Optimización del cálculo del fitness (métodos numéricos):
 $X * (A + D) = X * A + X * D$, $X * A$ precalculado y $X * D$ coste lineal.

Problemática estadística

- Normalmente se pretende minimizar un estimador estadístico.
- Puede incluir el determinante de la matriz de covarianzas (Akaike).
- La solución del problema de mínimos cuadrados se puede usar como punto de partida para optimizar el estimador con búsqueda local,
- o se puede optimizar directamente con el estimador usando metaheurísticas.

Diversidad de problemas

- Identificación de variables externas.
- Determinación de la amplitud de las dependencias temporales.
- Intervalo de realización de medias para evitar influencia de fluctuaciones.
- Variación del modelo con el tiempo: series temporales de modelos de series temporales.

Campos de aplicación

- Datos meteorológicos.
- Datos médicos, proyecto de Jose Juan López Espín.
- Datos económicos

Por ahora, con datos económicos aproximaciones satisfactorias, con datos médicos peores predicciones.

Alfonso L. Castaño, Javier Cuenca, Jose Matias Cutillas Lozano, Domingo Giménez, Jose-Juan López-Espín, Alberto Pérez-Bernabeu: Parallelism on Hybrid Metaheuristics for Vector Autoregression Models. HPCS 2018: 828-835

Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltente de Datos (DEA)
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos**
- 5 Sistemas Multicuerpo
- 6 Créditos

Problema de determinación de componentes

- Problema del Grupo de Polímeros del Departamento de Química-Física de la Universidad de Murcia.
<http://leonardo.inf.um.es/macromol/>
- Análisis de sedimentación de una solución compuesta, a partir de planteamiento de José García de la Torre. Analysis of Sedimentation experiments, 2014.
- Determinar la concentración de cada componente en el compuesto. Cada componente dado por su masa, M_k , y coeficiente de fricción, f_k .

Experimento

- Se centrifuga el compuesto. Medidas en distintos instantes de tiempo, t_j , $1 \leq j \leq n_t$, de una señal óptica en posiciones con distancias r_i , $1 \leq i \leq n_r$.
- La señal es función de la posición y del tiempo: $z(r, t)$.
- En instante inicial, con la solución en reposo, la señal tiene el mismo valor z_0 en todas las posiciones r , $z(r, 0) = z_0$.
- Solución con n_k componentes, cuya contribución a la señal es z_k , con $1 \leq k \leq n_k$.
- La señal z es aditiva en los componentes: $z = \sum_{k=1}^{n_k} z_k$.
- Cada z_k es de la posición r y el tiempo t .
- Para cada componente tenemos el peso molecular, M_k , y el coeficiente de fricción, f_k
- y queremos obtener la concentración del componente en la solución, y_k .

Determinación de la señal

- Se realizan varios experimentos, $1 \leq l \leq n_{exp}$
- que dependen de parámetros experimentales, e_l , y entre los que está la velocidad de rotación, ω_l .
- La señal de un experimento para una posición y un tiempo determinado es $z_{exp,l}(r_i, t_j; e_l)$, o simplificando $z_{exp,l}(r_i, t_j)$,
- y se obtiene de forma aditiva a partir de las señales de los componentes:

$$z_{exp,l}(r_i, t_j; e_l) = \sum_{k=1}^{n_k} z_k(r_i, t_j; M_k, f_k, y_k; e_l)$$

- y como la contribución a la señal de cada componente es lineal:

$$z_{exp,l}(r_i, t_j; e_l) = \sum_{k=1}^{n_k} (y_k \cdot z_k(r_i, t_j; M_k, f_k; e_l))$$

Simulación

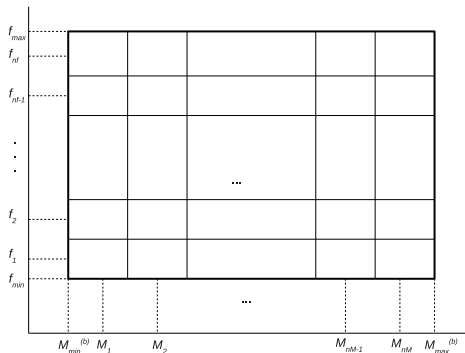
- Se dispone de procedimientos teóricos con los que estimar los valores de señal para cada experimento, en cada momento de tiempo y en cada posición, $z_{cal,l}(r_i, t_j; e_l)$.
- Obtener los valores de y_k con los que la diferencia entre los $z_{exp,l}$ y $z_{cal,l}$ se minimicen. Problema de mínimos cuadrados:

$$\Delta^2(p) = \frac{1}{n_{exp}} \frac{1}{n_r} \frac{1}{n_t} \sum_{l=1}^{n_{exp}} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \left(z_{exp,l}(r_i, t_j; e_l) - \sum_{k=1}^{n_k} y_k \cdot z_{k,cal,l}(r_i, t_j; e_l) \right)^2$$

- Sin conocer los z_{cal} ni los z_{exp} , podemos estudiar métodos de determinación de los y_k suponiendo valores de $z_{cal,l}$ determinados y valores de y_k , se generan con ellos los $z_{exp,l}$ que se obtendrían, se perturban, y se aplica el método de resolución del problema de mínimos cuadrados para determinar cuanto de lejos está la solución obtenida con la utilizada en la generación de los datos experimentales.

Malla de valores

- Se considera una distribución uniforme de los parámetros peso molecular y coeficiente de fricción.
- Los M en rango $[M_{min}, M_{max}]$, y los de f en $[f_{min}, f_{max}]$.
- Se discretizan con n_M y n_f valores:

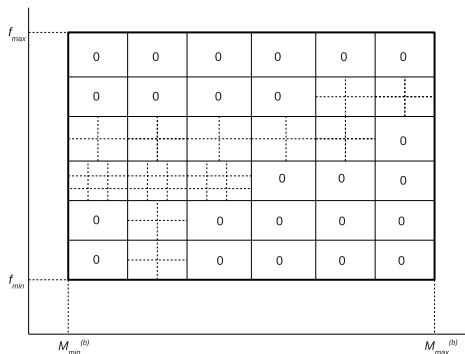


Parámetros del problema

- Típicamente datos de entre 1 y 5 experimentos,
- alrededor de 200 posiciones
- y 200 instantes de tiempo.
- Número de ecuaciones entre $4 \cdot 10^4$ y $2 \cdot 10^5$.
- Si se toman 20 valores de peso molecular y otros 20 para el coeficiente de fricción, el número de incógnitas es 200.

Refinamiento de la malla

- Se puede resolver con esos valores y descartar combinaciones con aportaciones al compuesto muy bajas (valores de y_k por debajo de un umbral)
- y resolver nuevos problemas definiendo mallas más finas en zonas del mallado donde la aportación sea mayor:



Solución numérica

- Se forma la matriz A obtenida con la aplicación de la ecuación para los $n = n_{exp} \cdot n_r \cdot n_t$ experimentos, posiciones en la celda que rota e instantes de tiempo.
- El número de incógnitas es $m = n_M \cdot n_f$.
- Tenemos $A \in R^{n \times m}$,
- y se plantea el problema a optimizar $\min_y \|Ay - b\|$, con b es el vector de señales obtenidas experimentalmente.
- El problema tenemos la restricción de que $0 \leq y_k \leq 1, \forall k / 1 \leq k \leq n_k$, y $\sum_{k=1}^{n_k} y_k = 1$.
- Es un problema de mínimos cuadrados no negativos, o NNLS (Non-Negative Least Squares), con variables acotadas y con una restricción lineal.
- En Matlab está la rutina `lsqlin`.
- Necesaria realización de rutinas paralelas y/o utilización de librerías paralelas

Aproximación metaheurística

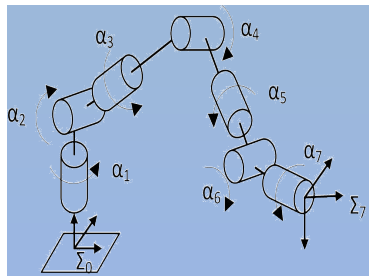
- Se buscan valores de y_k para el problema de mínimos cuadrados.
- Se generan individuos (valores de los y_k con $\sum_{i=1}^k y_k = 1$)
- la función de fitness es la diferencia entre los z_{exp} y los z_{cal}
- y se utilizan técnicas de combinación y mejora de elementos para ir mejorando el conjunto de individuos:
Algoritmos Genéticos, Scatter Search, Ascensión de colinas, Colonia de hormigas...
- Necesaria implementación de metaheurísticas paralelas y/o utilización de librerías paralelas de optimización

Contenidos

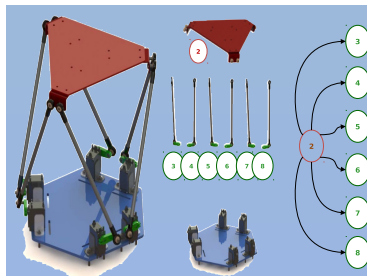
- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltante de Datos (DEA)
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo**
- 6 Créditos

Estructura con grupos de ecuaciones

Sistema multicuerpo



Plataforma de Stewart



José-Carlos Cano, Javier Cuenca, Domingo Giménez, Mariano Saura-Sánchez, Pablo Segado-Cabezas: A parallel simulator for multibody systems based on group equations. *The Journal of Supercomputing* 75(3): 1368-1381 (2019)

Simulador

High Performance Kinetic Simulator: Model Builder

Open model | New model | Delete model | Check | Save | Execute | DB | About | Model 1: Stewart

Groups (8) | Scenarios (2) | Routines (5) | Scripts (2) | Training scenario | Config

Select a Group: G4 | New Group | New Final Group | Generate trees

Group name: G4 | Delete Group | Clone group | Add a child

Routine: tres

Successors

Group Name	Delete	↑	↓
G8			

G1 | Add

F(x, ...)	Parameter	Matrix Name	Index
SOLVESYS	inMatrix	M1	3
SOLVESYS	inTritmindep	M2	3

Delete unused | Delete all

Matrix Name
M1
M2
M3
M4
M5
M6

Auto layout | Rotate | Zoom to Fit | Zoom in | Zoom out | Show scenarios

- Utilización del simulador para optimización de rutinas de álgebra lineal en entornos heterogéneos.
- Planteamiento de un **problema de optimización**: determinar posición inicial y parámetros de impulso para obtener una trayectoria deseada. El cálculo del fitness consiste en simular la trayectoria (métodos numéricos) y comparar con la deseada.

Contenidos

- 1 Aplicaciones en el grupo de CCPP
- 2 Análisis Envoltante de Datos (DEA)
- 3 Series temporales
- 4 Determinación de componentes por análisis de sedimentos
- 5 Sistemas Multicuerpo
- 6 Créditos

En orden de aparición

Martín González, Jose J. López-Espín, Juan Aparicio, Domingo Giménez, Jesús T. Pastor: Using Genetic Algorithms for Maximizing Technical Efficiency in Data Envelopment Analysis, The International Conference on Computational Science, Reykjavík, Iceland, 1-3 June, 2015

Martin Gonzalez-Rodríguez, Jose-Juan López-Espín, Juan Aparicio, Domingo Giménez, **El-Ghazali Talbi:** A parameterized scheme of metaheuristics with exact methods for determining the Principle of Least Action in Data Envelopment Analysis. CEC 2017: 588-595

Alfonso L. Castaño, Javier Cuenca, Jose Matias Cutillas Lozano, Domingo Giménez, Jose-Juan López-Espín, Alberto Pérez-Bernabeu: Parallelism on Hybrid Metaheuristics for Vector Autoregression Models. HPCS 2018: 828-835

José García de la Torre. Analysis of Sedimentation experiments, 2014

José-Carlos Cano, Javier Cuenca, Domingo Giménez, Mariano Saura-Sánchez, Pablo Segado-Cabezas: A parallel simulator for multibody systems based on group equations. The Journal of Supercomputing 75(3): 1368-1381 (2019)

Información adicional en <http://dis.um.es/~domingo/>:

- Metaheurísticas paralelas: optimización y aplicaciones, Máster de Computación Paralela y Distribuida, Technical University of Valencia, May 2019
- Computación de altas prestaciones, una herramienta en ayuda de la ciencia, presentation the festivity of San Alberto, November, 14, 2014